

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-069436

(43)Date of publication of application : 16.03.2001

(51)Int.Cl.

H04N 5/76

G06F 17/30

G10L 17/00

G10L 15/00

G11B 27/00

G11B 27/10

(21)Application number : 11-242555

(71)Applicant : NIPPON HOSO KYOKAI <NHK>

(22)Date of filing : 30.08.1999

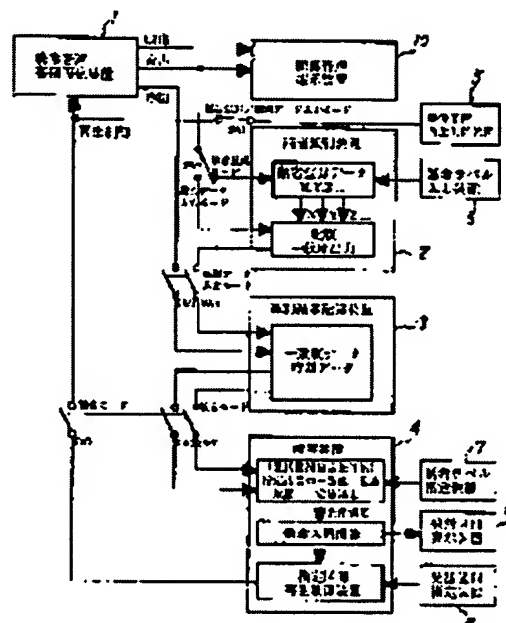
(72)Inventor : SUMIYOSHI HIDEKI

(54) SPEAKER INFORMATION RECORDING DEVICE AND RETRIEVING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To automatically record information such as speaker information corresponding to recorded video or audio by detecting matching degree for every speaker registered successively by being compared with the feature of the registered voice of the speaker and recording it on a recording medium together with a speaker name, program material or time information of a completed program.

SOLUTION: A video/audio storage and reproducing device 1 can simultaneously store and reproduce video and audio and is provided with a function for extracting the time information at the time of reproducing simultaneously with the stored video and audio and a function for reproducing video at designated time. A speaker identifying device 2 is provided with both a function for registering the speaker on the basis of an inputted voice and a function for obtaining a matching degree data expressed by numeric by comparing the registered voice of the speaker with the input voice. An identified result recording device 3 simultaneously records the coincidence degree data from the speaker identifying device 2 and time information at the time when these data are extracted.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-69436

(P2001-69436A)

(43) 公開日 平成13年3月16日 (2001.3.16)

| (51) Int.Cl. ⁷ | 識別記号 | F I | テーマコード* (参考) |
|-------------------------------------|------|---------------|-------------------|
| H 0 4 N 5/76 | | H 0 4 N 5/76 | B 5 B 0 7 5 |
| G 0 6 F 17/30 | | G 1 1 B 27/00 | 5 C 0 5 2 |
| G 1 0 L 17/00 | | 27/10 | 5 D 0 1 5 |
| 15/00 | | G 0 6 F 15/40 | 3 7 0 D 5 D 0 7 7 |
| G 1 1 B 27/00 | | 15/401 | 3 1 0 A 5 D 1 1 0 |
| 審査請求 未請求 請求項の数 3 O L (全 7 頁) 最終頁に続く | | | |

(21) 出願番号 特願平11-242555

(22) 出願日 平成11年8月30日 (1999.8.30)

(71) 出願人 000004352

日本放送協会

東京都渋谷区神南2丁目2番1号

(72) 発明者 住▲吉▼ 英樹

東京都世田谷区砦1丁目10番11号 日本放送協会 放送技術研究所内

(74) 代理人 100059258

弁理士 杉村 暁秀 (外2名)

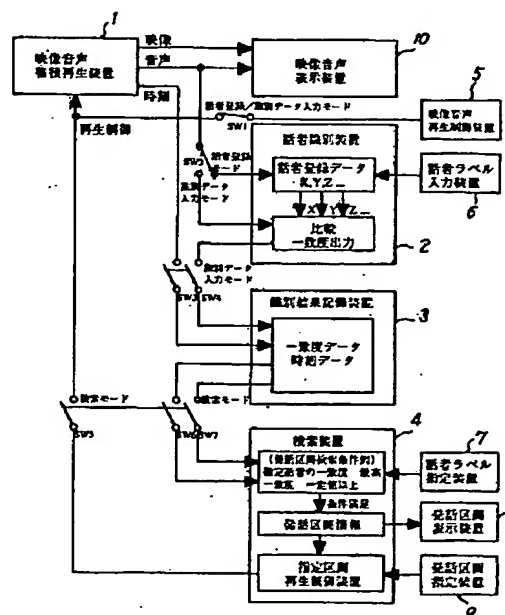
最終頁に続く

(54) 【発明の名称】 話者情報記録装置および検索装置

(57) 【要約】

【課題】 テレビジョン番組の制作過程において、撮影された映像素材に話者情報を付加する作業は、非常に時間がかかるとともに、集中力が要求される作業となっている。

【解決手段】 録画、録音テープ等に収録されている映像、音声の話者が誰であるのかの情報（話者情報）を、テープ中でその話者が発話している時刻とともに自動的に記録媒体に記録し、また、テープに収録されている複数話者のうちから特定の話者が発話している時間帯を検索し得るようにした。



【特許請求の範囲】

【請求項 1】 番組素材または完成した番組中の任意所望の話者の音声の特徴を抽出して当該話者を登録する手段、

前記番組素材または完成した番組、またはそれ以外の番組素材または完成した番組中で発話しているすべての音声の特徴を抽出しながら、前記登録された話者の音声の特徴と逐次比較して登録された話者毎の一致度を検出する手段、および該手段により検出された前記話者毎の一致度を、話者名および前記番組素材または前記完成した番組における時刻情報とともに記録媒体に記録する手段を少なくとも具えてなることを特徴とする話者情報記録装置。

【請求項 2】 請求項 1 記載の話者情報記録装置によって記録された前記話者毎の一致度、前記話者名および前記時刻情報を前記記録媒体から再生する手段、および該手段により再生された前記話者毎の一致度について、該一致度が前記登録された話者のうちで最も高い値を示し、かつ予め定めた値以上である場合、前記最も高い値を示す一致度に対応する登録された話者名を出力する手段を少なくとも具えてなることを特徴とする話者情報検索装置。

【請求項 3】 請求項 2 記載の話者情報検索装置において、該検索装置はさらに、前記登録された話者名を出力する手段から出力された話者名と検索しようとする話者名とがある時間帯において一致したとき、当該時間帯を、前記検索しようとする話者が前記番組素材または完成した番組、またはそれ以外の番組素材または完成した番組中で発話している時間帯であると判定する手段を具えてなることを特徴とする話者情報検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、収録された映像または音声（または、撮影された映像）に対して話者情報を収録（または、撮影）開始からの時間とともに記録媒体に記録する話者情報記録装置、および記録された記録媒体から特定の話者が発話している時間帯を検索する話者情報検索装置に関する。

【0002】

【従来の技術】テレビジョン番組の番組制作過程においては、放送時間の 10 倍～100 倍程度の時間数の映像素材を撮影しており、これら映像素材の蓄積、管理、検索に大きな労力がかかっている。放送番組を制作するには、これらの映像素材の中から番組で使用する映像を抜きだし、編集することで 1 つの番組に組み立てる。大量の映像素材の中から必要な映像を探すためには、撮影内容を情報（インデックス）として事前に記録しておくことが必要になる。

【0003】現在、この作業は完全に人手で行われてお

り、「どの映像素材（テープ番号など）」の、「先頭からどのくらいの位置（時間的位置）」に、「何が写っている」か、などの情報を番組制作者が VTR テープを再生して、映像を見ながらノートに書き起こし（記録）している。これは非常に時間がかかるとともに、集中力が要求される作業となっている。

【0004】また、上記のほか、完成番組の再利用を目的に映像内容情報を検索システムに登録することも行われているが、これら登録の作業に際しても、同様に人手で「どこに」、「何が写っている」かの記録を行う必要がある。

【0005】

【発明が解決しようとする課題】撮影内容を表現するに際しては、撮影された「人物」、「物体」、「場所」および「状況」などが用いられる。しかし、膨大な時間数にのぼる撮影された映像に対して、これら情報の記録を行うことはきわめて大きな作業負担となっている。

【0006】そのため、現在では番組制作者が映像、音声の収録された VTR テープを再生しながら書き起こしている「どの映像素材（テープ番号など）」の、「先頭からどのくらいの位置（時間的位置）」に、「何が写っている」という情報の記録の自動化が渴望されている。

【0007】本発明の目的は、収録された映像または音声（または、撮影された映像）に対して上述した種類の情報（特に、話者情報）の記録を自動化することにある。

【0008】

【課題を解決するための手段】上記目的を達成するために、本発明話者情報記録装置および検索装置は、録画、録音テープ等に収録されている映像、音声の話者が誰であるかの情報（話者情報）を、テープ中でその話者が発話している時刻（テープ先頭からの時間）とともに自動的に記録媒体に記録し、また、テープに収録されている複数話者のうちから特定の話者が発話している時間帯を検索し得るようにしたものである。

【0009】すなわち、本発明話者情報記録装置は、番組素材または完成した番組中の任意所望の話者の音声の特徴を抽出して当該話者を登録する手段、前記番組素材または完成した番組、またはそれ以外の番組素材または完成した番組中で発話しているすべての音声の特徴を抽出しながら、前記登録された話者の音声の特徴と逐次比較して登録された話者毎の一致度を検出する手段、および該手段により検出された前記話者毎の一致度を、話者名および前記番組素材または前記完成した番組における時刻情報とともに記録媒体に記録する手段を少なくとも具えてなることを特徴とするものである。

【0010】また、本発明話者情報検索装置は、話者情報記録装置によって記録された前記話者毎の一致度、前記話者名および前記時刻情報を前記記録媒体から再生す

る手段、および該手段により再生された前話者毎の一致度について、該一致度が前記登録された話者のうちで最も高い値を示し、かつ予め定めた値以上である場合、前記最も高い値を示す一致度に対応する登録された話者名を出力する手段を少なくとも具えてなることを特徴とするものである。

【0011】また、本発明話者情報検索装置は、さらに、前記登録された話者名を出力する手段から出力された話者名と検索しようとする話者名とがある時間帯において一致したとき、当該時間帯を、前記検索しようとする話者が前記番組素材または完成した番組、またはそれ以外の番組素材または完成した番組中で発話している時間帯であると判定する手段を具えてなることを特徴とするものである。

【0012】

【発明の実施の形態】以下に添付図面を参照し、発明の実施の形態に基づいて本発明を詳細に説明する。図1は、本発明話者情報記録装置および検索装置の一実施形態をブロック図にて示している。図1において、1は映像音声蓄積再生装置、2は話者識別装置、3は識別結果記録装置、4は検索装置、5は映像音声再生制御装置、6は話者ラベル入力装置、7は話者ラベル指定装置、8は発話区間表示装置、9は発話区間指定装置、および10は映像音声表示装置である。

【0013】以下に、符号1から4で示される本発明装置中の主要構成要素について順を追って説明する。まず、映像音声蓄積再生装置1は、映像と音声と同時に蓄積し、かつ再生することのできる装置である。この装置はまた、再生時の時刻情報を蓄積されている映像、音声と同時に取り出す機能と、指定された時刻の映像を再生する機能とを具えている。

【0014】話者識別装置2は、当該装置に入力される音声に基づいて話者を登録する機能（後述する話者登録モード参照）と、登録された話者の音声と入力音声とを比較して数値などで表される一致度データを得る機能（後述する識別データ入力モード参照）とを兼ね備えた装置である。

【0015】上記において、話者登録は、話者の音声の特徴をデータ化して登録することによって行われ、登録と識別（入力音声に登録された話者の音声と同一であるか否かの識別）は、特別な言葉ではなく自由に会話した自由音声を利用して行うことができる。

【0016】具体的には、話者の音声を一定時間（2分程度）入力し、音声データから特徴を抽出して話者データとして登録する（例えば、図1における話者登録データXとして）。登録された話者の音声を含む音声を、例えば、映像音声蓄積再生装置1から入力し、一定時間内（10秒程度）の音声データから登録時と同様に音声データの特徴を抽出する。登録話者の音声データとの類似度を数値で出力して一定値以上の場合にその登録話者が

発話しているものと判定することで、話者の識別を行うことができる。なお、本実施形態における話者識別装置2はコンピュータによって構成するが、音声データからの特徴抽出には話者認識ソフトウェアとして市販されているものを利用することができる。

【0017】識別結果記録装置3は、話者識別装置2からの一致度データとそのデータを取り出したときの時刻情報を（本実施形態においては、映像音声蓄積再生装置1から入力される）を同時に記録する（これについても、後述する識別データ入力モード参照）装置である。また、識別結果記録装置3は映像音声蓄積再生装置1の記録形態に応じて、テープ番号やファイル名といった、記録媒体を識別するための情報（媒体識別子）も同時に記録することができる。

【0018】また、検索装置4は、登録された話者の番号（後述する）や名前などを指定して、一致度データが一定の条件を満たしたときの時刻情報を取り出し（後述する検索モード参照）、さらに、必要に応じて映像音声蓄積再生装置1から映像、音声を映像音声表示装置10に入力して映像、音声の再生を行う装置である。

【0019】以下に、図1に示すスイッチSW1～SW7を用いて回路の接続替えを行い、話者登録、識別データ入力および検索を行う場合の回路動作について説明する。まず、話者登録の場合の回路接続を図2に示している。また、話者登録を行うモードのことを話者登録モードと呼ぶ。

【0020】話者登録モードにおいては、

- ・映像音声蓄積再生装置1に必要な番組素材の映像と音声を入力する（図示しない）、
 - ・スイッチSW1およびSW2をそれぞれ図示の話者登録モード側に切り替える、
 - ・映像音声再生制御装置5を操作し、登録したい話者が発話している音声を話者識別装置2に2分程度（長さは話者識別装置2に依存する）入力し、話者登録を行う、
 - ・登録時には、話者識別装置2に話者を特定する番号、名前などのラベルを話者ラベル入力装置6から入力する、
 - ・複数の話者を登録する場合には、上記の話者登録操作を話者の人数分だけ繰り返す、
- の操作を行う。

【0021】次に、識別データ入力の場合の回路接続を図3に示している。また、識別データ入力を行うモードのことを識別データ入力モードと呼ぶ。識別データ入力モードにおいては、

- ・スイッチSW1、SW2、SW3およびSW4をそれぞれ図示の識別データ入力モード側に切り替える、
- ・映像音声蓄積再生装置1を映像音声再生制御装置5により操作して映像、音声を再生し、そのうちの音声を話者識別装置2に入力する。このときに再生する映像、音声は、登録時に使用したものでも他の番組素材映像、音

声でもよい、

・話者識別装置 2 は、入力された音声とすでに登録されている話者データの音声の特徴を比較し、一致度を話者のラベルとともにデータとして出力し、その一致度データと話者のラベル、さらに映像音声蓄積再生装置 1 からの時刻情報を識別結果記録装置 3 に入力して記録する、
 ・映像が、複数の媒体（テープ、ディスクなど）や複数のファイルなどに分割されている場合は、それらの媒体識別子を付加して記録を繰り返す、の操作を行う。

【0022】次に、検索の場合の回路接続を図 4 に示している。また、検索を行うモードのことを検索モードと呼ぶ。

・蓄積された映像音声の中から、特定の話者が発話している部分を検索する場合に、スイッチ SW5、SW6 および SW7 をそれぞれ図示の検索モード側に切り替える。話者登録時に付加した番号、名前などのラベル情報を話者ラベル指定装置 7 から入力し、検索する話者を指定する、

・検索装置 4 は、識別結果記録装置 3 から一致度データ等の識別結果を読み出し、その中で「指定された話者の一致度データが一番高く、また一定値以上である部分」などの検索条件を満足する部分を見つけ、記録されている媒体識別子と時刻情報を取り出して発話区間表示装置 8 の画面に映し出して検索者に提示する、

・映像音声蓄積再生装置 1 に、発話区間指定装置 9 によって指定される発話区間に対応する部分の映像、音声が蓄積されているならば、検索装置 4 内の指定区間再生制御装置により制御して、媒体識別子と時刻情報を用いて発話区間の映像、音声を映像音声表示装置 10 に表示する、の操作を行う。

【0023】最後に、本発明話者情報記録装置および検索装置は、上述した実施形態に限られるものでなく、各種変形して実施することができる。以下に、上記以外の 5 つの想定される実施形態を列挙する。

・実施形態（その 1）

映像音声蓄積再生装置 1 と識別結果記録装置 3（図 1 参照）を一体化し、話者ラベル、一致度データを映像、音声と同期して記録するようにする。

・実施形態（その 2）

識別結果記録装置 3 に記録する情報が、話者ラベルと、
 「指定された話者の一致度データが一番高く、また一定値以上である部分」の条件を満足している時刻情報であ

る。

・実施形態（その 3）

実施形態（その 2）で記録する形式のデータを、映像音声蓄積再生装置 1 の記録内容管理情報として記録する。

・実施形態（その 4）

撮影装置（記録装置一体型テレビジョンカメラ装置など）に本発明装置を一体化し、撮影時に話者情報を映像、音声と同時に記録するようにする。

・実施形態（その 5）

10 映像音声蓄積再生装置 1 に音声情報のみ記録し、音声のみで構成する番組制作時に指定話者の検索などに用いる。

【0024】本発明話者情報記録装置および検索装置は、音声信号により話者の識別を行うようにしているので、ラジオなどの音声信号のみを取り扱う番組の素材管理、番組管理用に構成することも可能である。

【0025】

20 【発明の効果】本発明によれば、映像素材に対する情報を作成するにあたり、発話者の特定により、音声、および同時に撮影された映像への人物情報付与作業を自動化することができる。また、映像、音声に付与される発話者の情報は、媒体識別子、時刻情報とともに電子データとして蓄積されるので、発話者を指定することにより映像の検索が可能となる。

【図面の簡単な説明】

【図 1】 本発明話者情報記録装置および検索装置の一実施形態をブロック図にて示している。

【図 2】 話者登録の場合の回路接続を示している。

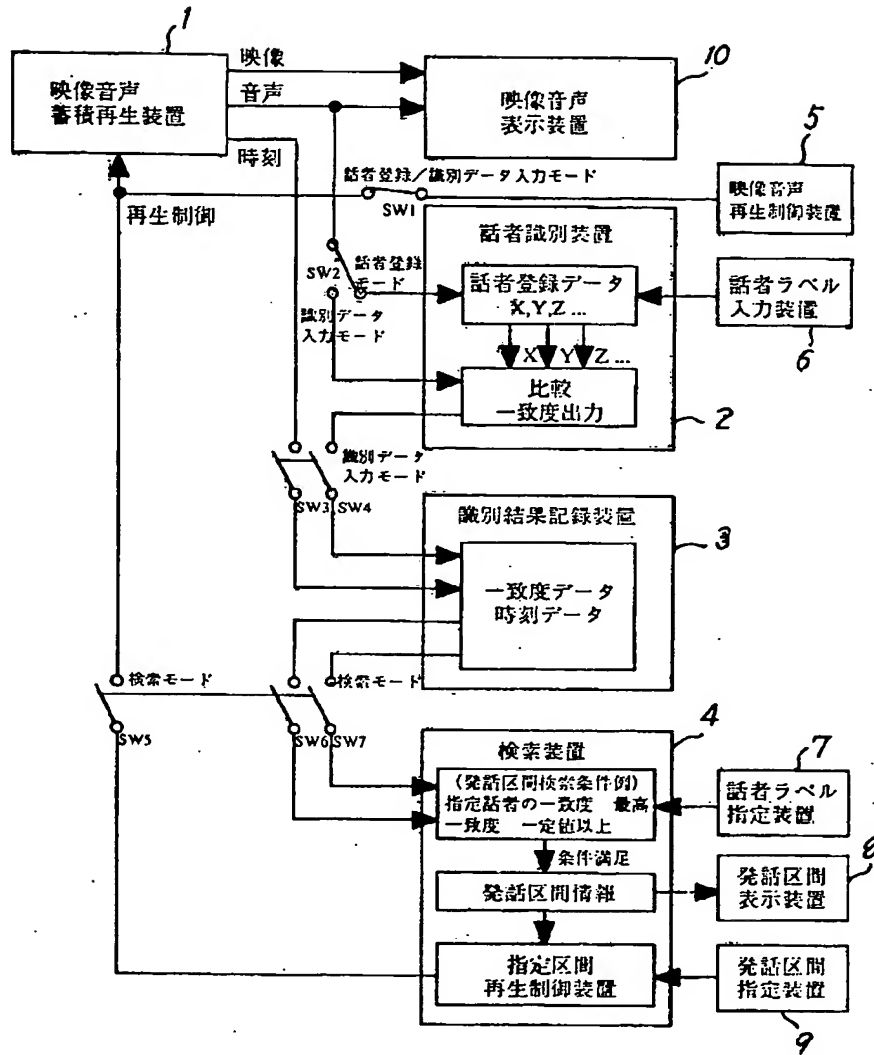
30 【図 3】 識別データ入力の場合の回路接続を示している。

【図 4】 検索の場合の回路接続を示している。

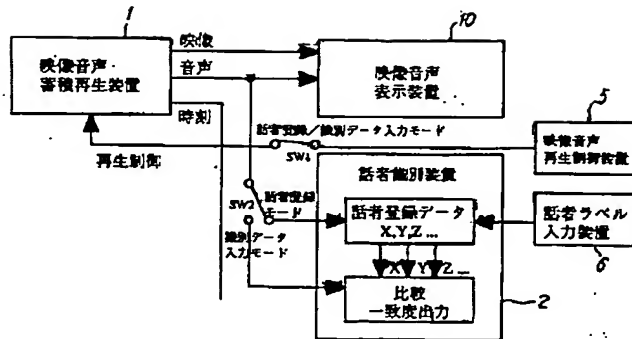
【符号の説明】

- 1 映像音声蓄積再生装置
- 2 話者識別装置
- 3 識別結果記録装置
- 4 検索装置
- 5 映像音声再生制御装置
- 6 話者ラベル入力装置
- 7 話者ラベル指定装置
- 8 発話区間表示装置
- 9 発話区間指定装置
- 10 映像音声表示装置

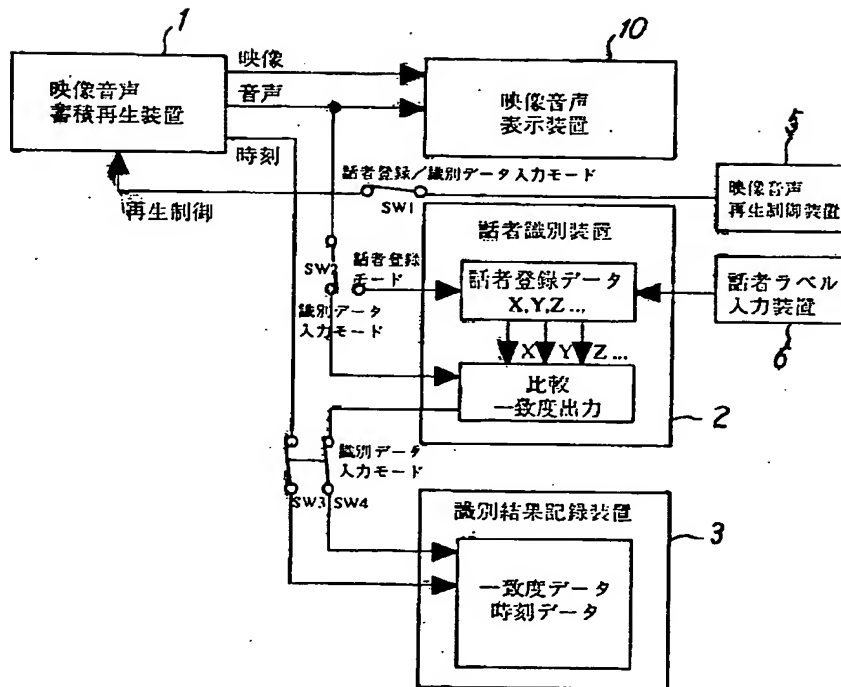
〔図1〕



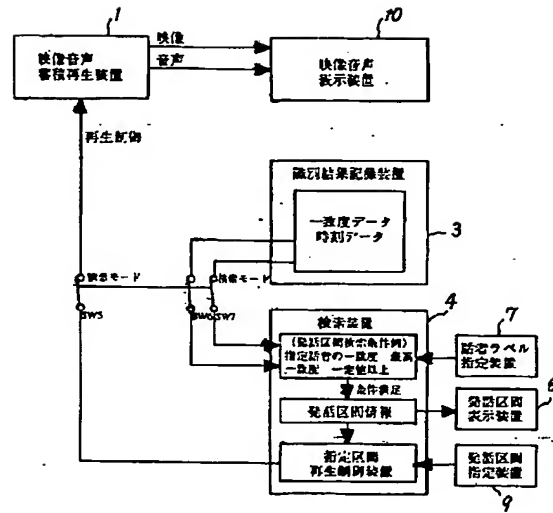
【図2】



【図3】



【図4】



フロントページの続き

(51)Int.Cl.⁷

G 1 1 B 27/10

識別記号

F I

G 1 0 L 3/00

G 1 1 B 27/00

27/10

ターマコード (参考)

5 4 5 A 9 A 0 0 1

5 5 1 G

A

A

F ターム (参考) 5B075 ND12 NK06 NK13 NR06 PQ02

UU34

5C052 AA01 AB02 AB10 AC08 DD04

5D015 AA02 AA03 KK02

5D077 AA02 BA08 CA02 CB07 DC22

DC32 EA13

5D110 AA02 AA27 DA04 DA12 DB17

DD11 DE05 EA06 EA07

9A001 BB01 BB03 BB04 FF03 HH15

HH16 HH17 HH23 HH28